UCRL-CONF-203283

Hierarchical Molecular Modelling with Ellipsoids

N. Max

March 31, 2004

Symposium on Point-Based Graphics
Zurich, Switzerland, Switzerland
June 2, 2004 through June 4, 2004

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Hierarchical Molecular Modelling with Ellipsoids

Nelson Max
Lawrence Livermore National Laboratory

Abstract

Protein and DNA structures are represented at varying levels of details using ellipsoidal RGBA textured splats. The splat texture at each level is generated by rendering its children in a hierarchical model, from a distribution of viewing directions, and averaging the result. For rendering, the ellipsoids to be used are chosen adaptively, depending on the distance to the viewpoint. This technique is applied to visualize DNA coiling around nucleosomes in chromosomes.

Categories and Subject Descriptors I.3.7 [Three-Dimensional Graphics and Realism]

1. Introduction

Macromolecular assemblies within a cell can be very complex, including multiple chains of protein, DNA, RNA, and other molecules. Level of detail simplification is useful to get an overall visualization at interactive rates. This paper describes hierarchical molecular simplification with ellipsoidal textured splats.

This is similar to the QSplat algorithm of Rusinkiewicz and Levoy [RL00], where a KD tree was used to construct a hierarchy of spheres to represent the surface of a solid object. But molecular assemblies already have a biochemically meaningful hierarchy, with amino acids built from atoms, protein chains built from amino acids, and multi-chain assemblies built from separate chains. The current system uses this natural hierarchy, uses ellipsoids instead of spheres to better approximate the shapes of the objects, and uses textures to approximate their colors and opacities.

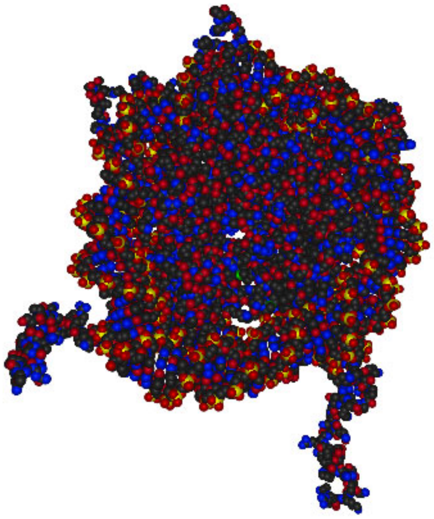

Figure 1: *Nucleosome with all non-hydrogen atom spheres.*

The ellipsoids are projected by the methods described in [ZPvB*02]. The main innovation in this paper is that the color/opacity texture for each object's splat, rather than a gaussian, is constructed from the average of rendered images of the object's children in the hierarchy, from a distribution of viewing directions.

2. Nucleosome structure

The nucleosome is a core assembly of eight histone protein chains, around which almost two full coils of DNA are wrapped, giving the disc-like particle shown in figure 1. These particles, connected by linker stretches of DNA, are arranged in a helical pattern to form a 30-nm chromatin fiber (see figure 8-13 of [ABL*94], and figure 2 here), which is itself further coiled when a chromosome condenses during cell division. This organized coiling keeps the very long DNA strands in a chromosome from getting tangled.

In the early 1980's I participated in the production of a dome-screen stereo computer animated film for the Fujitsu pavilion at Expo '85 in Tsukuba, Japan, which included a sequence showing DNA coiling into a condensed chromosome. (See [HLM*85] and [M85].) At this time, full structural information for the nucleosome was not known, so we just used spherical splats for the histone proteins. We used geometrical morphing for the DNA representation, from a ball-and-stick representation, to a "ladder" representation, ant then into a single thick line, as described in [M85]. In the lowest level of detail, the chromatin fiber was rendered with opaque cylindrical splats with shading and highlights, feathered into transparency at their ends, so that they could join into an approximation of a curved tube. (See figure 4 of [M85], also appearing as figure 1 of [M90].) This is similar to the way the current system renders the DNA double helix at a lower level of detail. The method of sorting the objects in depth and compositing them from back to front is also similar to the system presented here.

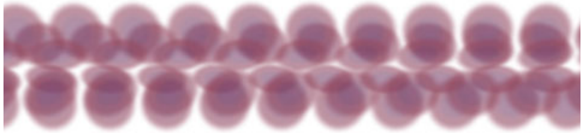

Figure 2: *A 30-nm chromatin fibre, rendered with a single splat per nucleosome.*

Since that time, a high resolution X-ray crystallography structure of the nucleosome with 147 base pairs of DNA has been solved [LMR*97], with atomic coordinates for all atoms except hydrogen, which is too light to diffract X-rays effectively. This structure was obtained from entry 1KX5 in the Protein Data Bank [PDB04]. The 12 base pair double helix of DNA from entry 119D was used as the linker to connect the nucleosomes. It was rotated and translated to continue one end of the nucleosome DNA. The union of these two structures has a total of 14,103 atoms. The individual chains are identified, as well as each amino acid in the proteins and each nucleotide in the DNA.

3. Level of detail Hierarchy

The first (highest) level of detail is the individual atoms shown in figure 1, which form the leaves of the hierarchy tree. The splats for the atoms use a separate RGBA texture for each atom type. So the data structure for an atom contains the center, radius, and texture square index into the larger texture shown in figure 3. The other nodes in the hierarchy tree at lower levels of detail are for ellipsoidal objects with a center, six floats for the upper triangular elements of the moment of inertia matrix, a bounding sphere radius for view frustum culling and level of detail choice, a texture square index, and a linked list of children.

The atom types appearing in protein and DNA are nitrogen, carbon, oxygen, sulphur, and phosphorus. I used two different shades for oxygen atoms: an orange-red for ones in the phosphate (PO4) groups in the DNA sugar-phosphate backbone, and a purpler shade for all the others. This distinguished the phosphate groups, and thus the DNA backbone.

The atom textures have an anti-aliased partially transparent profile curve, and a highlight in the center. All the object textures are blocks in the single 256 x 256 texture shown in figure 3. Counting across rows starting with position 1 at the upper left, the atom textures are in positions 24 through 29, in the order nitrogen, carbon, phosphate oxygen, sulfur, phosphorus, and non-phosphate oxygen.The atoms (and also the other ellipsoids) are sorted in depth, and composited from front to back, by applying the corresponding texture to a polygon surrounding the projection of the atom sphere. The radius for each atom type is the atom's Van der Walls radius, giving the standard intersecting sphere "space filling" molecular models. The intersection curves of these spheres are not generated by compositing disc-like planar texture splats, so the spheres pop unnaturally in front of each other as the viewpoint changes. Therefore I have added the option of rendering the front faces of an enclosing polyhedron, using the same texture. Latitude and longitude divisions are chosen at 45 ° intervals, dividing the visible half of a sphere with the North pole pointing towards the viewer into a triangle fan of 8 triangles, surrounded by a quad strip of 8 quads. This compromise gives only a rough approximation to the true intersection curve, but is quick to render. Since the polyhedron is circumscribed around the sphere, the shape of the texture map circle still gives a smooth profile to the sphere; only the intersection curve appears somewhat jagged when viewed up close. The lower levels of detail use softer edged ellipsoids of more similar colors, so the popping is less of a problem, and a single textured rectangle is used. Z-buffer testing is turned off, and visibility is determined only by the depth-sorted compositing order.

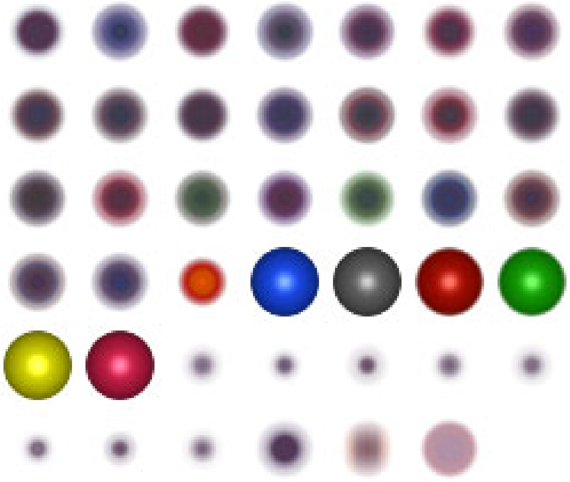

Figure 3: *Splat textures. In row order starting from the upper left are splats for 19 amino acids, 4 nucleosides, phosphate, 6 atoms, 8 protein chains, the union of all 8 protein chains, a cylinder for DNA, and the complete nucleosome.*

The next level of detail in the protein hierarchy is the amino acids, which are joined together to make a protein chain. There are 20 common amino acid types, only 19 of which appear in the nucleosome structure. Each type has a certain amount of flexibility, and can assume different shapes depending on its environment in the protein. So each instance of an amino acid has its own moment of inertia matrix, describing its shape as well as its orientation. The inverse of this matrix gives the quadratic form for the ellipsoid equation. But there is only one texture square for each of the amino acid types, constructed from one representative shape, giving the first 19 texture squares figure 3.

If a 3D ellipsoid is projected in perspective, the result is a 2D ellipse. The mathematics for the ellipsoid, its perspective projection, and its expansion for antialiasing, are given in [ZPvB*02]. The 2D ellipse is enclosed in a rectangle determined by its major and minor axes, and composited with the appropriate texture map. Finding the major and minor axes amounts to finding the eigenvectors of a quadratic form, which is relatively simple in 2D. Neither the inverse of a 3 by 3 matrix, nor 3D eigenvectors, are required.

For the ellipsoidal objects, a preprocessing step creates the texture squares in order, from higher level of detail to lower, by rendering its child objects, whose textures have already been determined, from multiple orthogonal views, and reading back the images from the RGBA frame buffer. On each image, the corresponding projected ellipse is conceptually superimposed. Its quadratic form gives the square of the Mahalanobis distance to the center, with contours of equal distance representing concentric ellipses. For each image, histograms of average color and opacity as a function of the Mahalanobis distance are constructed, and a weighted sum of these histograms determines the circularly symmetric texture square as a function of radius. The resulting texture gives an approximation to the average appearance of the child objects from any viewing angle, and its opacity gives the average coverage.

The sampled viewing directions are taken at equally spaced latitude and longitude angles on a unit sphere. The solid angles of the Voronoi regions of these samples are unequal, but are easily computed, and are used as weights in computing the weighted average of the histograms.

For DNA, the corresponding unit of structure is the nucleotide, consisting of a phosphate group attached to a nucleoside. The nucleoside, in turn, is made up of a sugar and a one of four bases, adenine, guanine, cytosine, and thymine. (See [H95] or [ABL*94].) The nucleotides are linked into a chain via their phosphates. In the DNA double helix, the phosphate-sugar linkages form two spiral backbones around the outside of the helix cylinder, and the bases, whose sequence determines the genetic code, stack inside, in planes approximately perpendicular to the cylinder axis.

I chose to separate out the phosphate groups in the object hierarchy, in order to delineate the spiral backbones. Since phosphorus is colored yellow and the phosphate oxygens are colored orange-red, this gives orange spherical splats for the phosphates. The four nucleosides are approximated by long flattened ellipsoids, so the base stacking is still visible in this level of detail. The four nucleoside textures are next after the amino acids in figure 3, followed by the orange phosphate texture. An image of the nucleosome at the level of detail showing amino acids, phosphates, and nucleosides, is shown in figure 4.

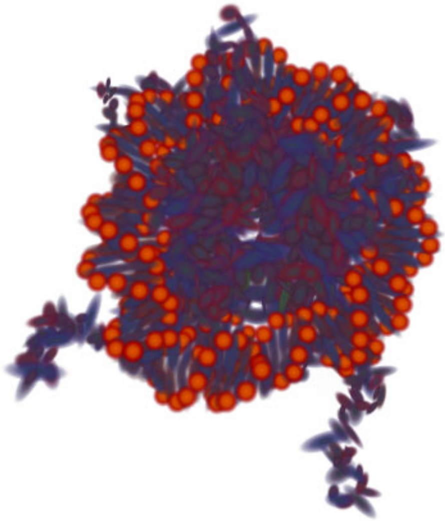

Figure 4: *The nucleosome at the level of detail showing amino acids, phosphates, and nucleosides.*

The next lower level of detail uses one ellipsoid for each of the eight amino acid chains in the nucleosome core. These are shown in the next 8 positions in the texture in figure 3. Since some of these chains are long and extended, they are not well approximated by ellipsoids, and diffuse density remains outside the central core of the texture. (I truncated the texture at a smaller radius to remove some of this excess density.) The following texture in figure 3 is for the whole nucleosome core, the union of these eight chains.

For the DNA at this level of detail, I used a structure consisting of 10 base pairs. Since this many base pairs gives a full turn of the DNA, the major axis of the resulting ellipsoid will point along the axis of the double helix, rather than becoming tilted because the center of gravity of the individual base pairs is displaced away from the axis.

In order to give the DNA helix the appearance of a bent cylindrical tube, cylindrical symmetry was assumed. The color and opacity histograms were created as a function of the projection, onto the minor axis direction of the 2D ellipsoid, of the vector from the texture square center to the pixel. The resulting 2D texture of vertical stripes was feathered to zero at the top and bottom, using a circular window, so that the splats of the DNA would fit together to form an approximation to a curved cylindrical tube. This texture is shown as the next to last one in figure 3. Note the orange edges, caused by the projection of the orange phosphate spheres concentrated at the profile of the double helix. A nucleosome at this level of detail is shown in figure 5. In addition to the 8 histone protein chains, a 50% dense image of the full nucleosome core was used to partially fill in the gaps between the protein chain ellipsoids.

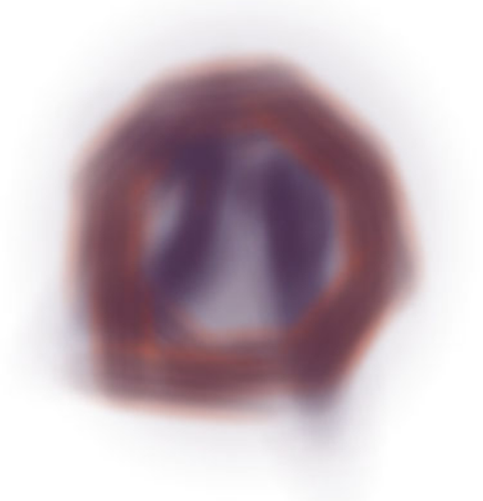

Figure 5: *The nucleosome at the level of detail showing protein chains and DNA cylinders.*

At the lowest level of detail, the whole nucleosome is represented by a single ellipsoidal splat, whose texture, the final one in figure 3, was constructed by viewing the structure of figure 5 from multiple directions.

Figure 2, showing a chain of 80 nucleosomes, was drawn using this level of detail. Instead of generating a root node with 80 copies of a rotated and translated nucleosome, the nodes selected for rendering were inserted in the list of objects to be sorted and rendered by calling the hierarchy traversal procedure 80 times with different model-view matrices. This meant that it was not necessary to store moment of inertial matrices for all the instances of all ellipsoids in the model, nor 4 by 4 sub-positioning matrices for each child object with respect to its parent object. The effect of the model-view rotation on the moment of inertia matrix is explained in [ZPvb*02].

4. Hierarchy traversal

The rendering step begins by constructing the list of objects to be sorted. In order to get a smooth “cross dissolve” transition between the levels of detail, this list may contain both an object and its children, each with opacity weights for the cross dissolve. These weights are used on the polygon vertices, and multiply the colors and opacities from the texture map. The opacity of an object at an internal hierarchy node begins to fade out in favor of its children when the ratio of its boundary sphere radius to its distance from the viewpoint is greater than a threshold T1, and completely fades out when this ratio reaches a larger threshold T2. When the hierarchy is being traversed, the opacity of a node’s parent is also known, and the opacity actually used for a node is the minimum of its own opacity, and one minus the opacity of its parent. This makes each ellipsoid visible only for an appropriate interval of viewing distances. Pseudocode for the hierarchy traversal thus uses the recursive routine putSort:

```
putSort(node *n, float ap) {
  if(sphere(n->center, n->radius) is outside
    the view frustum) return;
  float an = max(0., (min 1.,
    (T2 - n->radius/n->depth)/(T2 - T1)));
  if(an > 0.) putList(n, min(an, 1. - ap));
  if(an < 1.)
    for all children *c of *n
      putSort(c, an);
  }
```

Before the list of objects and associated opacities produced by the putList calls is sorted in depth, an extra displacement is added to the depth of all atoms. This causes the Z-buffered atom spheres to be composited before their parent splats, making the cross dissolve appear more natural. After the quicksort is completed, the spheres or ellipsoids are projected in back to front order, and the corresponding textured rectangles are composited.

The first submitted animation ran at an average speed of 9 frames a second on a 800 MHz dual Pentium4 Xeon PC with an nVidia 5900FXUltra graphics card. It shows an exponential zoom towards the chromatin fiber, starting with figure 2 and ending with a close-up of individual atom spheres. The second animation, moving along the fiber, ran at 2 frames a second, since there was much more occluded geometry. Occlusion culling could probably speed this up.

5. Future work

The sphere intersection curves are jagged when viewed in close-up, due to the rough polyhedral approximation. To fix this, the number of faces in the polyhedron could be increased as the spheres come closer. Alternatively, the sphere depth could be calculated at each pixel using a fragment program or displacement texture.

Protein domains are compact subsets of a protein chain, which may have evolved independently, and later joined by gene rearrangement. They can be more closely approximated by ellipsoids than can they whole protein chain, and could be an intermediate level of detail between the amino acids and the whole chain. This domain structure could be added to the system for files in which it is available (it was not for the 1KX5 nucleosome), or an automated domain subdivision algorithm could be written.

Acknowledgements

This work was performed under the auspices of the U. S. Department of Energy by Lawrence Livermore National Laboratory under contract W-7405-ENG-48. I thank Mike Pique of the Scripps Research Institute for giving me a multicolored physical “rapid prototype” model of the nucleosome, which helped me understand its structure.

References

[ABL*94] Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.: *Molecular biology of the cell; third edition*, Garland Publishing, 1994.

[HLM*85] Hirata, F., Max, N., Lerner, D., Okada, T.: Computer graphics: the FACOM M-380 creates an OMNIMAX, stereo movie, *Fujitsu 36*, 2 (1985) pp. 151 - 165. (in Japanese).

[LMR*97] Luger, K., Maeder, A., Richmond, R., Sargent, D., Richmond, T.: Crystal structure of the nucleosome core particle at 2.8 A resolution, *Nature 389* (1997) pp. 251 - 260.

[M85] Max, N.: DNA animation, from atom to chromosome, *Journal of Molecular Graphics 3*, 2 (1985) pp. 69 - 71 (with correct figures in the next issue).

[M90] Max, N., Cone-spheres, *Computer Graphics 24*, 4 (Siggraph 1990) pp. 59 - 62.

[PDB04] http://www.rcsb.org/pdb/

[RL00] Rusinkiewicz, S., Levoy, M.: A multiresolution point rendering system for large meshes. *ACM Computer Graphics Proceedings, Annual Conference Series, 2000*, pp. 343 - 352.

[ZPvB*02] Zwicker, M., Pfister, H., van Baar, J., Gross, M.; EWA splatting, *IEEE Transactions on Visualization and Computer Graphics 8*, 3 (2002), pp. 223 - 238.

[H95] Hallick,R.: http://www.blc.arizona.edu/Molecular_Graphics/DNA_Structure/DNA_Tutorial.HTML (1995)